Many interspecific chromosomal introgressions are highly prevalent in Holarctic Saccharomyces uvarum strains found in human-related fermentations

Warren Albertin, Maria Chernova, Pascal Durrens, Erwan Guichoux, David James Sherman, Isabelle Masneuf-Pomarede, Philippe Marullo

MANY INTERSPECIFIC CHROMOSOMAL INTROGRESSIONS ARE HIGHLY PREVALENT IN HOLARCTIC SACCHAROMYCES UVARUM STRAINS FOUND IN HUMAN-RELATED FERMENTATIONS

Albertin Warren 1,2# , Chernova Maria 1# , Durrens Pascal 3,4, Guichou Erwan 5, Sherman David 3,4, Masneuf-Pomarede Isabelle 1,6, Philippe Marullo 1,7*

1. Univ. Bordeaux, ISVV, Unité de recherche OEnologie EA 4577, USC 1366 INRA, Bordeaux INP, 33140 Villenave d'Ornon, France
2. ENSCBP, Bordeaux INP, 33600 Pessac, France
3. CNRS UMR 5800, Univ. Bordeaux, 33405 Talence, France
4. Inria Bordeaux Sud-Ouest, joint team Pleiade Inria/INRA/CNRS, 33405 Talence, France
5. INRA, UMR1202 Biodiversité Gènes et Ecosystèmes, Plateforme Génomique, Cestas 33610, France
6. Bordeaux Sciences Agro, 33170 Gradignan, France
7. Biolaffort, 33100 Bordeaux, France

"#" authors contributed equally to this paper "*" author to whom to correspond Philippe Marullo, Unité de recherche OEnologie EA 4577, ISVV, 210 chemin de Leysotte, 33140 Villenave d'Ornon, France. +33557575864 Email: philippe.-marullo@u-bordeaux.fr

Abstract

In the last two decades, the massive genome sequencing of strains belonging to the Saccharomyces genus revealed the complex reticulated evolution of this group. Among the various evolutionary mechanisms described, the introgression of large chromosomal regions resulting from interspecific hybridization has recently shed light on the Saccharomyces uvarum species. In this work we provide the de novo assembled genomes of four S. uvarum strains presenting more than 712 kb of introgressed loci inherited from both S. eubayanus and S. kudriavzevii species. In order to study the prevalence of such introgressions in a large population, we designed multiplexed PCR- markers able to survey the inheritance of 8 chromosomal regions. Our data confirm that introgressions are

widely disseminated in Holarctic S. uvarum population and are more frequently found in strains isolated from human-related fermentations. According to the origin of the strains (nature, cider or wine related process) some loci are over-represented suggesting their positive selection by human activity. Except one locus located on chromosome 7, the introgressions present a low level of heterozygozity similar to that observed for nine neutral markers (microsatellites). Finally, most of these loci tested showed an expected Mendelian segregation after meiosis and can recombined with their chromosomal counterpart in S. uvarum.

Embank accession number: SAMN07178572 to SAMN07178575 Key words : Wine, cider, introgression, Saccharomyces eubayanus, Saccharomyces uvarum, MassARRAY.

Introduction

Deciphering genome evolution in yeast has gained momentum these last two decades with the release of hundreds of genome sequence of several yeast species including Saccharomyces cerevisiae 1–5, its sister Saccharomyces species 2,6–9 and other ascomycetes 10–13. Several prominent mechanisms of genomic evolution have been described, among them interspecific hybridization14,15, reticulated evolution 16, aneuploidization 1,17, recent or ancient polyploidization events 9,12, large chromosomal duplication or more limited gene duplication 18,19, horizontal transfer 20. These mechanisms are usually so closely intertwined that it is difficult to determine which one(s) is(are) cause(s) or consequence(s), but regardless they have drastically shaped yeast genome along evolution; see 21–23 for extensive reviews. Introgression is one such evolutionary mechanism. It has been described so far in different yeast species 24,25, but has been particularly addressed within the Saccharomyces genus 26–29.

Introgression is defined as the transfer of large or more limited genetic information from one species to another, and results in mosaic genomes, whose formal characterisation has long been complicated due to the lack of appropriate molecular tools 30. Introgression can be the result of interspecific hybridization followed by the extensive loss of one parental genome, either through repeated backcross with one parental species or through mis- segregation of the hybrid at meiosis. In any case, the preferential loss of one parental genome (except for the

introgressed regions) may allow the restoration of meiotic fertility, and subsequent successful sexual reproduction. Alternatively, Horizontal Gene Transfer (HGT) may account for the advent of introgressed regions, as it is in the case of in Saccharomyces cerevisiae where Zygosaccharomyces bailii 20 and Torulaspora microellipsoides 31 introgressions have been identified in the wine yeasts group. The mechanism of HTG could be mediated by episomal replication 32. Introgression has been largely reported as a mechanisms driving rapid adaptive evolution in yeast 33 and other eukaryotes 34,35 including human 36, animals 37, or plants 38. It is therefore not surprising that introgression has been frequently associated with domestication in all eukaryotic kingdoms 34,39,40.

Saccharomyces uvarum is a striking example of a yeast species whose genome is strongly shaped by introgressed regions 27. S. uvarum shares partially overlapping ecological niches with S. cerevisiae: both are strongly related to human-driven fermentation, but S. uvarum is more psychrotrophic and thus is more frequently associated with low-temperature processes: cider-making and wine-making in northern –cooler– French vineyards for example 41,42,43,44. Isolates from natural environments (insect, plant, soil) have been also described 45,46 . In 2014, Almeida et al. performed comparative genomics of 54 S. uvarum strains 27. Unexpectedly, 21 of these strains presented introgressions, the number of introgressed regions and their size being highly variable among isolates (up to 900 kb of introgressed regions). These introgressions derived mostly from the sister species S. eubayanus and possibly resulted from a few inter-specific hybridization events followed by chromosomal rearrangements and the extensive loss of most of the S. eubayanus genome, excepting the introgressed regions. Almeida et al. pointed out several interesting features: i- all strains displaying introgressed regions originated from the Northern hemisphere; ii- within the Holarctic population, S. eubayanus introgressions seemed to be more prevalent in strains associated with human activities (and largely absent from wild isolates); iii- those introgressions were significantly enriched in genes involved in nitrogen and sulphite metabolism. These results feed the hypothesis that selective pressures in anthropic environments have promoted the selection of multiple introgressions in Holarctic domesticated isolates.

In this paper, we developed tools to rapidly assess the presence of introgressed regions in a large population of S. uvarum isolates (104 strains). Since introgressed regions were absent from Southern hemisphere isolates, we decided to focus on Holarctic isolates from natural, cider and wine environments. We confirm that the overall number of introgressed regions is significantly higher in

cider-associated strains compared to wild strains, and is furthermore higher in wine isolates. However, only a subset of the introgressed regions were found to be overrepresented in anthropic activities and their number and quality varied between cider- and wine-making processes. Finally, we investigated the meiotic segregation of those introgressions in F1-hybrid progenies, demonstrating their Mendelian inheritance.

Materials and methods

Yeast strains used and culture media

All the strains used in this study are described in Table 1. The genomes of four strains of Saccharomyces uvarum (U1, U2, U3 and U4) have been sequenced in this work. The strains so named were obtained by tetrad microdissection47 and are monosporic clones of the strains PM12, PJP3, BR6-2 and RC4-15, respectively. Their genomic sequences (short reads) have been previously released 27. A collection of one hundred and four strains of S. uvarum sampled from various isolation substrates (grape/wine, nature, cider and others fermented beverages) was also genotyped. All of these strains were isolated in the North hemisphere and could be considered to belong to the Holarctic group of S. uvarum 27. Furthermore, a few interspecific hybrids (CBS 3008, CBS 425, CBS 1480, CID1) were genotyped. Finally, two sets of meiotic progeny clones of S. uvarum F1-hybrids carrying different introgressions were also obtained by tetrad microdissection. The F1-hybrids used, UU23 and UU34, were previously obtained from haploid derivatives of U2, U3 and U4 47. In order to set up the genotyping method of introgressions in S. eubayanus, the strains belonging to S. kudriavzevii (ZP542), S. cerevisiae (VL3) and S. eubayanus (CBS 12357) were used. All strains were usually grown at 24°C in YPD medium containing 1% yeast extract (Difco Laboratories, Detroit, MI), 1% Bacto peptone (Difco), and 6% glucose, supplemented or not with 2% agar. Sporulation was induced in ACK medium (1%, potassium acetate, 2% agar) for 3 days at 24°C after an overnight preculture on YPD medium.

Genome assembly of four S. uvarum strains

The genomic sequences of strains U1, U2, U3 and U4 were obtained by combining both Illumina Pair End and Mate Pair datasets. Briefly, genomic DNA was extracted from a saturated culture of 100 ml under anaerobic condition (YPD) using the genomic tip-100 kit (Qiagen, Courtaboeuf, FRANCE). Paired-end and 2.5 kb Mate pair Illumina libraries were prepared according to manufacturer protocols (Genomic DNA Sample Preparation) from sonicated genomic DNA.

Sequencing was performed on Illumina Genome Analyzer IIx (Illumina, CA) with a read length of 54 pb by the Genomic and Transcriptomic facility of Bordeaux, FRANCE. A mapping dataset was obtained by mapping reads on the reference genome the S. uvarum CBS 7001 7 using the stampy program. Variant calling was performed by mapping short reads to the reference genome using stampy 48 followed by samtools 49,50. SNPs were called using samtools mpileup with mapping quality ≥ 30 and base quality ≥ 20, and varFilter depth ≥ 10. Single amino-acid polymorphisms (SAPs) were identified using snpEff 51, requiring quality QUAL ≥ 30 and genotype GEN[*].GQ ≥ 20.. A de novo assembly was then carried out from an initial set of 80x single-read combined with 180x paired-end sequences from 2500 +/- 250 bp inserts. Intitial contigs from GAIIx reads were assembled using Mira 3.2.1 52 with 8 passes. They were oriented and joined into scaffolds with paired-end sequences as follows. To anchor initial contigs into the paired-end assembly, they were fragmented into 45x 160 bp libraries using simLibrary 1.3 then into overlapping reads by simNGS 1.6 53 to simulate the AllPaths-LG sequencing protocol. These fragment reads were combined with the paired-end reads and reassembled using AllPaths-LG 54. The supercontigs for the 4 strains were deposited on GenBank database with the following BioProject ID: PRJNA388544. The genomes of the strains U1 to U4 are registered with the accession number SAMN07178572 to SAMN07178575, the genomes were not annotated. Whole-genome synteny was computed using Sibelia 55 using the "loose" parameter, pairwise between CBS7001 and strains U1–U4. A total of 1484 blocks longer than 5 kbp were found, with a median length of 45 kbp. Selected syntenies are shown in Figure S1 for U1 (CBS7001 chr.1, chr.10, cha.11), U2 (CBS7001 chr.1, chr.7, chr.13, chr.16), and U4 (CBS7001 chr.1, chr.8, chr.13), illustrating both coverage and collinearity.

Introgression genotyping

Rapid DNA extraction.

The genomic DNA of S. uvarum isolates and monosporic clones were quickly extracted in 96-wells microplate format using a customized LiAc-SDS protocol. Basically, 5.106 cells were pelleted in a PCR microplate and incubated with 50µL of 200mM LiAc/1% SDS at 70°C for 5 minutes. Genomic DNA was then extracted by mixing cell lysates with 150µL of pure ethanol and vortexed for 15 seconds. After a brief centrifugation (5 min, 4400 rpm) the supernatant was removed and the pellet washed with 70% ethanol. Genomic DNA was then solubilized in 200µL of milliQ water at 60°C for 5 minutes. After a brief centrifugation, cell de-

bris were pelleted and 150µL of supernatant containing genomic DNA was recovered in a new microplate. The genomic DNA was then analyzed by MassARRAY genotyping.

MassARRAY genotyping

Initially, 20 sequences located in the 8 introgressed regions were screened, corresponding to 74 polymorphic sites, including SNPs and INDELs. Candidate markers were submitted for assay design using the MassARRAY Assay Design version 4.0.0.2 (Agena Biosciences, Hamburg, Germany). T o circumvent the high polymorphism in each sequence (2-8 polymorphisms within 103-151bp), we decreased the allowed PCR primer length to 16 bases, reduced the minimum peak separation to 10Da and extended the mass array window to between 3000 and 10000Da. One multiplex of 30 polymorphisms was selected (Table S2), covering 16 out of the 20 sequences tested. We used 15 ng of DNA for genotyping with the MassARRAY iPLEX platform (Agena Bioscience) following manufacturer's instructions. Raw data analyses were performed using Typer Viewer v 4.0.26.75 (Agena Bioscience). We filtered out monomorphic SNPs, loci with weak or ambiguous signal (loci displaying more than three genotypic clusters or unclear cluster separation). The markers showed mean amplification rates of 95.5% (84.9% - 100%).

Genetic and statistical analyses

The genotypes of a subpanel of 72 S. uvarum strains were obtained from a previous genetic analysis using 9 microsatellite markers 56. Expected and observed heterozygosity were calculated from Hardy-Weinberg equilibrium using the ade4 package (R). Chi2 tests were performed to assess whether the proportion of heterozygous individuals was higher for introgressed markers compared to microsatellite ones (statistical test of the presence/absence of the introgressed markers). Chi2 test was applicable as all groups displayed >10 individuals as recommended by Cochran 57. A non-parametric statistical test (Kruskal-Wallis) was used to determine whether the strains of the different groups presented significantly different number of introgressed markers using R package agricolae.

The subpanel of 72 S. uvarum strains was then used to draw dendrogram trees using either microsatellite data or introgressed markers. Microsatellite tree was built using Bruvo's distance and NJ clustering (poppr package, R). Introgression tree was built using Euclidean distance and Ward's clustering. The genetic distance was estimated using the Haldane relation d=-1/2ln(1-2r), were r is the recombination rate.

Results and Discussion

Genome sequences of four monosporic clones of S. uvarum strains

The four monosporic clones U1 to U4 were obtained by tetrad microdissection respectively from strains PM12, PJP3, BR6 and RC4-15 and have been previously sequenced by a Pair end strategy 27. In order to release an improved genome quality, an additional sequence dataset was obtained with a 2.5 kb-mate pair approach (see methods). Using both data sets, the de novo assembly delivered nearly 50 scaffolds for the strains U2, U3 and U4. For some chromosomes, the assembled scaffolds correspond to an entire chromosome (Figure S1). The assembly of U1 is more fragmented than U2–U4, likely due to the poor quality of the mate pair library. Although not completely finished, the scaffolds released will contribute to enrich the genomic databases.

In the present study, we focused our attention on some genomic regions showing a strong SNP polymorphism density (higher than 5% of divergence) respect to the reference genome (CBS 7001). This high polymorphic rate contrasts with the relative low SNP polymorphism found for the remaining part of the genome that varies between 1.92 (U2) to 2.24 (U1) SNP/kb according to the strain (Figure 1A). The high polymorphic regions encompass 712 kb and are located in 8 Saccharomyces uvarum chromosomes (chr. 2, 4, 6, 7, 9, 13, 14, 16) corresponding to the interspecific introgressions from S. eubayanus and S. kudriavzevii described by Almeida et al. (2014). Except two large regions in chromosome 12 and 15 (detected for the strains DBVPG 7787 and 148.01 respectively), the four clones sequenced showed almost all of the interspecific introgressions described until now for this species. For each genome a blast analysis confirmed that all the introgressed loci belong to distinct scaffolds confirming that these regions were not physically linked.

Design of multiplexed PCR experiment for tracking S. eubayanus and S. kudriavzevii introgressions in a large set of S. uvarum strains.

To confirm the inheritance of these introgressions, 20 species-specific PCR markers covering the eight introgressions were designed. Each locus was covered by at least one marker and, for the larger ones, by few markers spaced every ~30 kb. The specificity of each marker was confirmed by using as templates the DNA of the strains CBS7001 (S. uvarum), ZP542 (S. kudriavzevii (European)), CBS 12357 (S. eubayanus) and VL3 (S. cerevisae). As expected, the locus located on the chromosome 13 (13_17) was amplified with the DNA of the strain ZP542 (S. kudriavzevii). All of the other loci were positively amplified us-

ing the strain CBS 12357 (S. eubayanus) but are not amplified by other reference strains of S. uvarum, S kudriavzevii and S. cerevisiae (data not shown). For all the markers, the strains U1 to U4 showed the allele inheritance predicted by the genomic sequence. The names, positions, and relative inheritance of these 20 PCR-markers were given in the Table 2.

In order to readily track these interspecific introgressions within a large set of S. uvarum strains, a high-throughput PCR screening was then developed. We used the MassARRAY technology, which allows genotyping up to 48 SNP in a single multiplexed reaction 58. Due to the very divergent sequence between the S. uvarum genome and the introgressed regions, only 16 loci of the 20 designed were positively multiplexed; each of the 8 chromosomes was covered by at least one marker. The Figure 1B shows the relative position of the MassARRAY markers on the S. uvarum CBS 7001 map.

Prevalence of introgression in strains associated to alcoholic fermentation.

The prevalence of the 16 MassARRAY markers was evaluated in a population of 104 holarctic S. uvarum strains isolated from different substrates: 13 isolates from nature, 60 strains from grape or wine, 29 from cider or fruits (except grapes) and 2 isolates of unknown origin. In addition, four interspecific hybrids CBS 3008, CBS 425, CBS 1480 and CID1, the S. eubayanus type strain (CBS 12357) as well as the fully homozygous strains U1 to U4 were genotyped. The whole dataset is represented on Figure 2. Only 5 S. uvarum strains (CBS 7001, CECT 10192, ZP1021, ZP554 and ZP556) displayed no introgressed markers, confirming the high prevalence (95%) of introgressed regions in Holarctic S.uvarum population. All of them belong to the 'nature' group have a very limited number of introgressed loci (only the markers 13_17 and 7_65). Although, the number of strains from the 'nature" group is limited in this study, this result confirm that most of the introgressions described are rare for such strains. The highest number of introgressed markers was 8 for strain TB95VIC3 (grape-wine group) and many strains have more than 5 introgressed markers. To determine whether the number of introgressions was significantly different depending on the substrate origin, we computed the average number of introgressed markers per group (Figure 2). Overall, strains from 'nature' displayed a mean of 0.61 introgressed markers, while strains from cider-fruit and grape-wine possessed 2.24 and 4.48 introgressed markers respectively. A Kruskal-W allis test indicated that the

over- representation of introgressed markers in both anthropic groups was significant compared to wild strains, and furthermore that grape-wine strains had the higher number of introgressed markers.

All of the introgressed regions derived from S. eubayanus species (chromosomes 2, 4, 6, 7, 9, 14, 16) were not or were poorly detected within the 'nature' population. Thus, we tested whether each marker was over-represented in cider-fruit and/or grape-wine groups compared to nature one (Chi2 test α=0.05, Figure 2). Grape-wine strains displayed a significant over-representation of 6 markers distributed over 3 chromosomes (4, 6 and 14). By contrast, cider-fruit isolates displayed only one overrepresented marker, located on chromosome 2. For the cider group the allele frequency was 3.7 fold higher than for the wine group (0.31 vs 0.083). One possible explanation of this enrichment could be due to the presence of the ASP1 gene encoding the cytosolic L-asparaginase (type I) and required for asparagine anabolism 59. Asparagine is the most abundant amino-acid in most apple juices (10-30mg/100ml apple juice) 60,61, while grape juices usually display 100 fold lower asparagine concentration 61. Interestingly, when L-asparagine is a major nitrogen source, the activity of L-asparaginase strongly impacts yeast growth as well as acetic acid production 62 which are important traits in both cider and wine industry.

The introgression located on chromosome 13 and derived from S. kudriavzevii showed an atypical inheritance and was the unique introgression harbouring a relatively high frequency in 'nature' group (allele frequency 0.29). Nonetheless, the marker 13_17 was still significantly over-represented in both cider- and wine-related populations and represented by far the most frequent allelic form. The relative high frequency of this S. kudriavzevii region in S. uvarum natural isolates might be explained by the fact that European S. kudriavzevii and S. uvarum shared the same biotope (bark tree) and temperature optimums (cold regions) 45. This environmental proximity might have promoted hybridization and/or horizontal transfer events.

Two additional introgression regions (not screened in this work) have been identified in only two 'nature' isolates: DBVPG 7787 (chr. 12) and 148.01 (chr. 15) 27. Our method could be applied to large nature isolate to test if these regions are more frequently found in natural populations and might confer any adaptation to wild habitat conditions. However their low frequency (each found only 2 times in 54 genomes) seems in contraction with any positive selection.

Finally, we tested whether the introgression patterns could be used as a proxy for genetic distance between S. uvarum strains. Among the 104 strains genotyped, a subpanel of 72 has been previously genotyped by using 9 microsatellite markers 56. Two dendrograms were built using either microsatellite or introgression data (Figure 3). The trees were not completely congruent, except for the most distant group (called "A" in the microsatellite tree) that globally is well conserved in the introgression tree. However, most of these strains have unique geographical and source origins (fermented grape juice, Sancerre, France) and might be strongly similar clonal variants. Therefore, further experiments are needed to increase the S. uvarum collection tested in order to have a more precise idea of the relationship between genetic diversity, geographical origin and possible domestication events.

"Genetic behavior" of introgressed loci

Previous analysis reported very low levels of heterozygosity in S.uvarum using microsatellite genotyping, probably as a consequence of a high selfing rate (>95%) 56. In order to test whether introgression and microsatellite displayed similar pattern regarding heterozygosity levels, we computed the observed and expected heterozygosity (from Hardy- Weinberg equilibrium) for both set of markers (Figure 4). For all markers, observed heterozygosity is around 10-fold lower than expected, in agreement with a high selfing rate. Expected heterozygosity is higher for microsatellite markers compared to introgression markers, probably as the consequences of the increased number of alleles for microsatellites. Observed heterozygosity ranged from 0 to 10%, the mostly heterozygous locus being 7_65. Interestingly, Almeida et al. (2014) discussed the possible selective advantage of chromosome 7 introgression, as it contains FZF1 gene involved in sulphite resistance 63,63,64 and ZRT1 gene that presents traces of balancing selection 66. In this work, we show that chromosome 7 introgression is not significantly overrepresented in cider- and wine-making processes compared to natural ones, in apparent contradiction with any selective advantage. This result underlines the difficulty to drawn correlation between functionnal genetics and presence/absence of particular alleles in limited populations. Interestingly, it has to be noted that chromosomes 7's introgression displays the higher level of heterozygosity (>10%), which is significantly higher than the proportion of observed heterozygosity within microsatellites (Chi2 test, α=0.05). Such observation raises the hypothesis of a possible heterozygous advantage.

Finally, we investigated the segregation of five introgressed regions (chr. 2, 4, 6, 14 and 15) by analyzing their inheritance in the meiotic progeny of two S. uvarum F1-hybrids. The hybrids UU23 and UU34 have been obtained by crossing haploid derivatives of the strain U3 with U2 and U4, respectively 47. The germination rate of each hybrid is close to 50 % and few complete tetrads were obtained in both cases (Table S4). All of the spore clones (UU23=73 and UU34=48) were genotyped by MassARRAY for the 10 markers covering the five introgressions (Table 3). As expected, most of the markers showed a Mendelian segregation and the five complete tetrads dissected displayed a 2:2 segregation (data not shown). For the introgression of chromosome 4, a slight but significant enrichment for the eubayanus allele was found (khi2, α=0.05). This result may indicate a trend toward positive selection of the S. eubayanus allele, which is also suggested by the strong frequency (52%) of this introgressed region in the wine group. For two loci, few recombination events were observed (chr 6: 1/73 within 9 kb; chr 14 6/48 within 305 kb) between S. uvarum and S. eubayanus alleles. The maximal ratio between genetic and physical distance for the two loci ranged between 0.05 to 0.17 cM/kb. Although lower than the average ratio observed in Saccharomyces cerevisiae (0.33 cM/kb), the rare crossing overs observed demonstrated that these interspecific regions have been successfully incorporated in the meiotic machinery of S. uvarum despite their high genetic divergence with the S. uvarum genome. The selection of appropriate spore clones of UU23 and UU34 containing S. eubayanus markers and their successive mating would result in the construction of strains presenting all the introgressed regions for chromosome 2, 4, 6, 14 and 15. By crossing such strains with selected U1 spore clones, most of the S. eubayanus introgressions should be grouped in the same hybrid in two crosses, offering new perspectives for studying whether those introgressions may confer a selective advantage and/or a phenotype of interest. Indeed, breaking the linkage disequilibrium existing within S. eubayanus alleles would be efficient for addressing the effect of introgression on phenotypes. The development of MassARRAY markers allowing the genotyping of numerous spore clones in a short time paves the way for quantitative genetics programs that are very efficient in yeast 67.

Conclusion

In this work, we show that 95% of Holarctic isolates of S. uvarum harbour introgressions where the number and the size of the introgressed regions depend on the strains. We confirmed that anthropic isolates significantly possess more in-

trogressions than wild strains. In addition, we showed that only one introgressed region is overrepresented for cider-making environment, and up to 3 regions for wine-related process. Interestingly, Almeida et al. (2014) reported that strains from the Northern hemisphere showed remarkably low diversity across their genomes compared to Southern hemisphere isolates, while previous microsatellite analysis failed to detect a significant clustering based on substrate origin 56. This quite low genetic diversity contrasts with the relative high phenotypic variability found for technological traits 68. This contradiction suggests that interspecific introgressions found among Holarctic S. uvarum strains could be the most important source of genetic, and by extent of phenotypic variability. The high throughput genotyping method developed here paves the way for studying the impact of these regions on the phenotypic variability of S uvarum strains.

Acknowledgements

Genome sequencing presented in the present publication were performed at the Genomic and Transcriptomic Facility of Bordeaux (grants from the Conseil Regional d'Aquitaine nu20030304002FA and 20040305003FA. PM received a grant from Conseil Regional d'Aquitaine (SAGESSE). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Gallone, B. et al. Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts. Cell 166, 1397–1410.e16 (2016).

2. Liti, G. et al. Population genomics of domestic and wild yeasts. Nature 458, 337–341 (2009).

3. Goffeau, A. et al. Life with 6000 genes. Science 274, 546, 563–7 (1996).

4. Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A. & Schmidt, S. A. Whole Genome Comparison Reveals High Levels of Inbreeding and Strain Redundancy Across the Spectrum of Commercial Wine Strains of Saccharomyces cerevisiae. G3 (Bethesda). 6, 957–71 (2016).

5. Wang, Q. M., Liu, W. Q., Liti, G., Wang, S. A. & Bai, F. Y. Surprisingly diverged populations of Saccharomyces cerevisiae in natural environments remote from human activity. Mol. Ecol. 21, 5404–5417 (2012).

6. Hittinger, C. T. et al. Remarkably ancient balanced polymorphisms in a multi-locus gene network. Nature 464, 54–58 (2010).

7. Scannell, D. R. et al. The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. G3: Genes|Genomes|Genetics 1, 11–25 (2011).

8. Nakao, Y. et al. Genome sequence of the lager brewing yeast, an interspecies hybrid. DNA Res. 16, 115–129 (2009).

9. Libkind, D. et al. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proc. Natl. Acad. Sci. U. S. A. 108, 14539–14544 (2011).

10. Souciet, J.-L. et al. Comparative genomics of protoploid Saccharomycetaceae. Genome Res. 19, 1696–1709 (2009).

11. Dujon, B. et al. Genome evolution in yeasts. Nature 430, 35–44 (2004).

12. Wong, S., Butler, G. & Wolfe, K. H. Gene order evolution and paleopolyploidy in hemiascomycete yeasts. Proc. Natl. Acad. Sci. U. S. A. 99, 9272–9277 (2002).

13. Borneman, A. R. et al. Insights into the Dekkera bruxellensis Genomic Landscape: Comparative Genomics Reveals Variations in Ploidy and Nutrient Utilisation Potential amongst Wine Isolates. PLoS Genet. 10, e1004161 (2014).

14. Dunn, B., Richter, C., Kvitek, D. J., Pugh, T. & Sherlock, G. Analysis of the Saccharomyces cerevisiae pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. Genome Res. 22, 908–24 (2012).

15. Leducq, J.-B. et al. Speciation driven by hybridization and chromosomal plasticity in a wild yeast. Nat. Microbiol. 1, 15003 (2016).

16. Peris, D. et al. Population structure and reticulate evolution of Saccharomyces eubayanus and its lager-brewing hybrids. Mol Ecol 23, 2031–2045 (2014).

17. Bond, U., Neal, C., Donnelly, D. & James, T. C. Aneuploidy and copy number breakpoints in the genome of lager yeasts mapped by microarray hybridisation. Curr Genet 45, 360–370 (2004).

18. Steenwyk, J. & Rokas, A. Extensive Copy Number Variation in Fermentation-Related Genes Among Saccharomyces cerevisiae Wine Strains. G3: Genes|Genomes|Genetics g3.117.040105 (2017). doi:10.1534/g3.117.040105

19. Fares, M. A., Keane, O. M., Toft, C., Carretero-Paulet, L. & Jones, G. W. The Roles of Whole-Genome and Small-Scale Duplications in the Functional Specialization of Saccharomyces cerevisiae Genes. PLoS Genet 9, e1003176 (2013).

20. Novo, M. et al. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. Proc. Natl. Acad. Sci. 106, 16333–16338 (2009).

21. Liti, G. & Louis, E. J. Yeast evolution and comparative genomics. Annu Rev Microbiol 59, 135–153 (2005).

22. Dujon, B. Yeast evolutionary genomics. Nat Rev Genet 11, 512–524 (2010).

23. Albertin, W. & Marullo, P. Polyploidy in fungi: evolution after whole-genome duplication. Proc. R. Soc. B Biol. Sci. 279, 2497–2509 (2012).

24. Kavanaugh, L. A., Fraser, J. A. & Dietrich, F. S. Recent evolution of the human pathogen Cryptococcus neoformans by intervarietal transfer of a 14-gene fragment. Mol Biol Evol 23, 1879–1890 (2006).

25. Mallet, S. et al. Insights into the life cycle of yeasts from the CTG clade revealed by the analysis of the Millerozyma (Pichia) farinosa species complex. PLoS One 7, e35842 (2012).

26. Liti, G., Barton, D. B. H. & Louis, E. J. Sequence diversity, reproductive isolation and species concepts in saccharomyces. Genetics 174, 839–850 (2006).

27. Almeida, P. et al. A Gondwanan imprint on global diversity and domestication of wine and cider yeast Saccharomyces uvarum. Nat. Commun. 5, 4044 (2014).

28. Muller, L. A. H. & McCusker, J. H. A multispecies-based taxonomic microarray reveals interspecies hybridization and introgression in Saccharomyces cerevisiae. FEMS Yeast Res 9, 143–152 (2009).

29. Naumova, E. S., Naumov, G. I., Masneuf-Pomarède, I., Aigle, M. & Dubourdieu, D. Molecular genetic study of introgression betweenSaccharomyces bayanus andS. cerevisiae. Yeast 22, 1099–1115 (2005).

30. Morales, L. & Dujon, B. Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. Microbiol. Mol. Biol. Rev. 76, 721–39 (2012).

31. Marsit, S. et al. Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. Mol. Biol. Evol. msv057- (2015). doi:10.1093/molbev/msv057

32. Galeote, V. et al. Amplification of a Zygosaccharomyces bailii DNA segment in wine yeast genomes by extrachromosomal circular DNA formation. PLoS One 6, e17872 (2011).

33. Dunn, B. et al. Recurrent Rearrangement during Adaptive Evolution in an Interspecific Yeast Hybrid Suggests a Model for Rapid Introgression. PLoS Genet 9, e1003366 (2013).

34. Ropars, J. et al. Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi. Curr. Biol. 25, 2562–2569 (2015).

35. Arnold, M. L. & Martin, N. H. Adaptation by introgression. J Biol 8, 82 (2009).

36. Huerta-Sanchez, E. et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512, 194–197 (2014).

37. Fitzpatrick, B. M. et al. Rapid fixation of non-native alleles revealed by genome-wide SNP analysis of hybrid tiger salamanders. BMC Evol Biol 9, 176 (2009).

38. Martin, N. H., Bouck, A. C. & Arnold, M. L. Detecting adaptive trait introgression between Iris fulva and I. brevicaulis in highly selective field conditions. Genetics 172, 2481–2489 (2006).

39. Giuffra, E. et al. The origin of the domestic pig: independent domestication and subsequent introgression. Genetics 154, 1785–1791 (2000).

40. Zhao, K. et al. Genomic diversity and introgression in O. sativa reveal the impact of domestication and breeding on the rice genome. PLoS One 5, e10780 (2010).

41. Masneuf-Pomarede, I., Bely, M., Marullo, P. & Albertin, W. The Genetics of Non- conventional Wine Yeasts: Current Knowledge and Future Challenges. Front. Microbiol. 6, (2016).

42. Tosi, E., Azzolini, M., Guzzo, F. & Zapparoli, G. Evidence of different fermentation behaviours of two indigenous strains of Saccharomyces cerevisiae and Saccharomyces uvarum isolated from Amarone wine. J. Appl. Microbiol. 107, 210– 218 (2009).

43. Demuyter, C., Lollier, M., Legras, J.-L. & Le Jeune, C. Predominance of Saccharomyces uvarum during spontaneous alcoholic fermentation, for three consecutive years, in an Alsatian winery. J. Appl. Microbiol. 97, 1140–1148 (2004).

44. Naumov, G. I. et al. Genetic identification of Saccharomyces bayanus var. uvarum, a cider-fermenting yeast. Int. J. Food Microbiol. 65, 163–71 (2001).

45. Sampaio, J. P. & Gonçalves, P. Natural Populations of Saccharomyces kudriavzevii in Portugal Are Associated with Oak Bark and Are Sympatric with Cerevisiae and S Paradoxus/. Appl. Environ. Microbiol. 74, 2144–2152 (2008).

46. Boynton, P. J. & Greig, D. The ecology and evolution of non-domesticated Saccharomyces species. Yeast 31, 449–62 (2014).

47. da Silva, T. et al. Hybridization within Saccharomyces Genus Results in Homoeostasis and Phenotypic Novelty in Winemaking Conditions. PLoS One 10, e0123834 (2015).

48. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21, 936–9 (2011).

49. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–9 (2009).

50. Danecek, P., Auton, A., Abecasis, G. & Albers, C. The variant call format and VCFtools. (2011).

51. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 6, 80–92 (2012).

52. Chevreux, B., Wetter, T. & Suhai, S. Genome sequence assembly using trace signals and additional sequence information. Ger. Conf. Bioinforma. (1999).

53. Massingham, T. & Goldman, N. simNGS and simLibrary–software for simulating next- gen sequencing data. (2012).

54. Gnerre, S., MacCallum, I. & Przybylski, D. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. (2011).

55. Minkin, I., Pham, H., Starostina, E., Vyahhi, N. & Pham, S. C-Sibelia: an easy-to-use and highly accurate tool for bacterial genome comparison. F1000Research 2, (2013).

56. Masneuf-Pomarede, I. et al. Microsatellite analysis of Saccharomyces uvarum diversity. FEMS Yeast Res. 16, (2016).

57. Cochran, W. The [chi-squared] test of goodness of fit. Ann. Math. Stat. 25, 315–345 (1952).

58. Gabriel, S., Ziaugra, L. & Tabbaa, D. SNP genotyping using the sequenom massARRAY iPLEX Platform. Current Protocols in Human Genetics (2009).

59. Dunlop, P. C., Meyer, G. M., Ban, D. & Roon, R. J. Characterization of two forms of asparaginase in Saccharomyces cerevisiae. J. Biol. Chem. 253, 1297–1304 (1978).

60. Burroughs, L. F. The amino-acids of apple juices and ciders. J. Sci. Food Agric. 8, 122–131 (1957).

61. Dizy, M., Martín-Alvarez, P. J., Cabezudo, M. D. & Carmen Polo, M. Grape, apple and pineapple juice characterisation and detection of mixtures. J. Sci. Food Agric. 60, 47– 53 (1992).

62. Marullo, P. et al. Single QTL mapping and nucleotide-level resolution of a physiologic trait in wine Saccharomyces cerevisiae strains. FEMS Yeast Res. 7, 941–952 (2007).

63. Avram, D., Leid, M. & Bakalinsky, A. T. Fzf1p of Saccharomyces cerevisiae is a positive regulator of SSU1 transcription and its first zinc finger region is required for DNA binding. Yeast 15, 473–480 (1999).

64. Park, S. K., Boulton, R. B. & Noble, a. C. Formation of hydrogen sulfide and glutathione during fermentation of white grape musts. Am. J. Enol. Vitic. 51, 91–97 (2000).

65. Engle, E. K. & Fay, J. C. Divergence of the yeast transcription factor FZF1 affects sulfite resistance. PLoS Genet. 8, (2012).

66. Coi, A. L. et al. Genomic signatures of adaptation to wine biological aging conditions in biofilm-forming flor yeasts. Mol. Ecol. (2017). doi:10.1111/mec.14053

67. Liti, G. & Louis, E. J. Advances in Quantitative Trait Analysis in Yeast. PLoS Genet. 8, (2012).

68. Masneuf-Pomarède, I., Bely, M., Marullo, P., Lonvaud-Funel, A. & Dubourdieu, D. Reassessment of phenotypic traits for Saccharomyces bayanus var. uvarum wine yeast strains. Int. J. Food Microbiol. 139, 79–86 (2010).

Conflict of interest : PM is an unpaid member of BIOLAFFORT group developing yeast strains starters for winemaking.

Tables and Figures

- **Table 1. Yeast strains used.**

- **Table 2. S. eubayanus and S. kudriavzevii introgressions detected by genome sequencing and confirmed by PCR.**

- **Table 3. Segregation analysis and recombination frequency of S. eubayanus introgressed loci.**

- **Figure 1. SNP scanning of four S. uvarum genomes defined eight interspecific introgressions tracked by MassARRAY genotyping.** Panel A. The number of SNP per kb relative to the S uvarum reference (CBS 7001) genome was given for the sequenced strains U1 to U4 color-coded according to the key. The eight genomic regions, having a SNP divergence higher than 5% and longer than 9 kb, are shaded in grey and are located on the chromosomes 2, 4, 6, 7, 9, 13, 14, and 15. Panel B. The names and positions of 20 markers designed are shown on the genetic map of the CBS 7001 reference genome. Dark blue and grey dots represented the markers multiplexed or not by MassARRAY. Finally, the names and positions of the nine microsatellites markers 56 used for calculating heterozygosity are shown by dark red dots.

- **Figure 2. Detection of introgressed 16 markers in 113 strains of S. uvarum and related species.** For each strain and each marker, a grey square indicates the presence of the S. uvarum-like allele, while a colored square indicates introgressions. Heterozygosity are represented by grey/colored triangle, and missing data by white square. For S. uvarum only, Chi2 test were performed to assess whether the introgressions were over- or under-represented depending on the substrate origin (nature [13 strains], cider-fruit [29 strains], grape-wine [60 strains]). Colored stars indicate significant distribution differences (α=0.05, Benjamini-Hochberg correction for multiple testing), and 'c' or 'w' indicate whether the introgressed markers are over represented for 'cider-fruit' and/or 'grape-wine' groups compared to nature one. The number of introgressed markers harboured by the S. uvarum of different substrate origin was calculated (bottom-left graph), and was found significantly different (Kruskal-Wallis, α =0.05, different letters indicates different means).

- **Figure 3. Dendrogram trees from microsatellite and introgressed markers.** 72 S. uvarum strains were genotyped for both set of markers and were used. The microsatellite tree was built using Bruvo's distance and NJ

clustering. The three main groups (A, B, C) were then reported on the introgression tree, built using Euclidean distance and Ward's clustering.

- **Figure 4. Observed and expected heterozygosity for introgressed and microsatellite markers.** Expected heterozygosity was calculated from Hardy-Weinberg equilibrium using the ade4 package (R), using only S. uvarum strains (nature, cider-fruit, grape-wine and unknown groups). Microsatellite data were extracted from 56. 'Mean Hobs' stands for mean observed heterozygosity calculated from microsatellite markers only. Chi2 tests were performed to assess whether the proportion of heterozygous individuals was higher for introgressed markers compared to those for microsatellites, only 7_65 markers was significant (α=0.05).